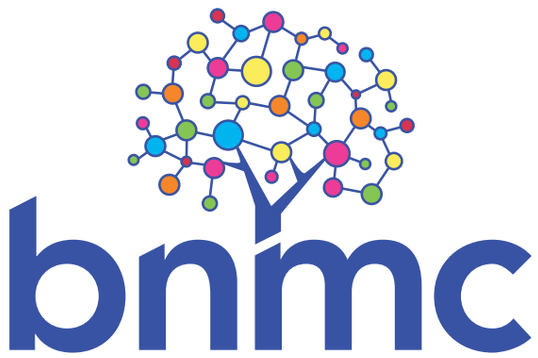
SBML, BioModels.net, and SBGN

Michael Hucka

Co-director—Biological Network Modeling Center (BNMC), Beckman Institute
Senior Research Fellow—Control and Dynamical Systems

California Institute of Technology
Pasadena, California, USA



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML background

SBML background

- Conviction that computational modeling becoming crucial
  - Mechanistic modeling and dynamical simulation enables **quantitative hypothesis testing**

SBML background

- Conviction that computational modeling becoming crucial
  - Mechanistic modeling and dynamical simulation enables **quantitative hypothesis testing**
- Not a new idea—dates to 1940's if not earlier

SBML background

- Conviction that computational modeling becoming crucial
  - Mechanistic modeling and dynamical simulation enables **quantitative hypothesis testing**
- Not a new idea—dates to 1940's if not earlier
- Today software tool support is better than ever

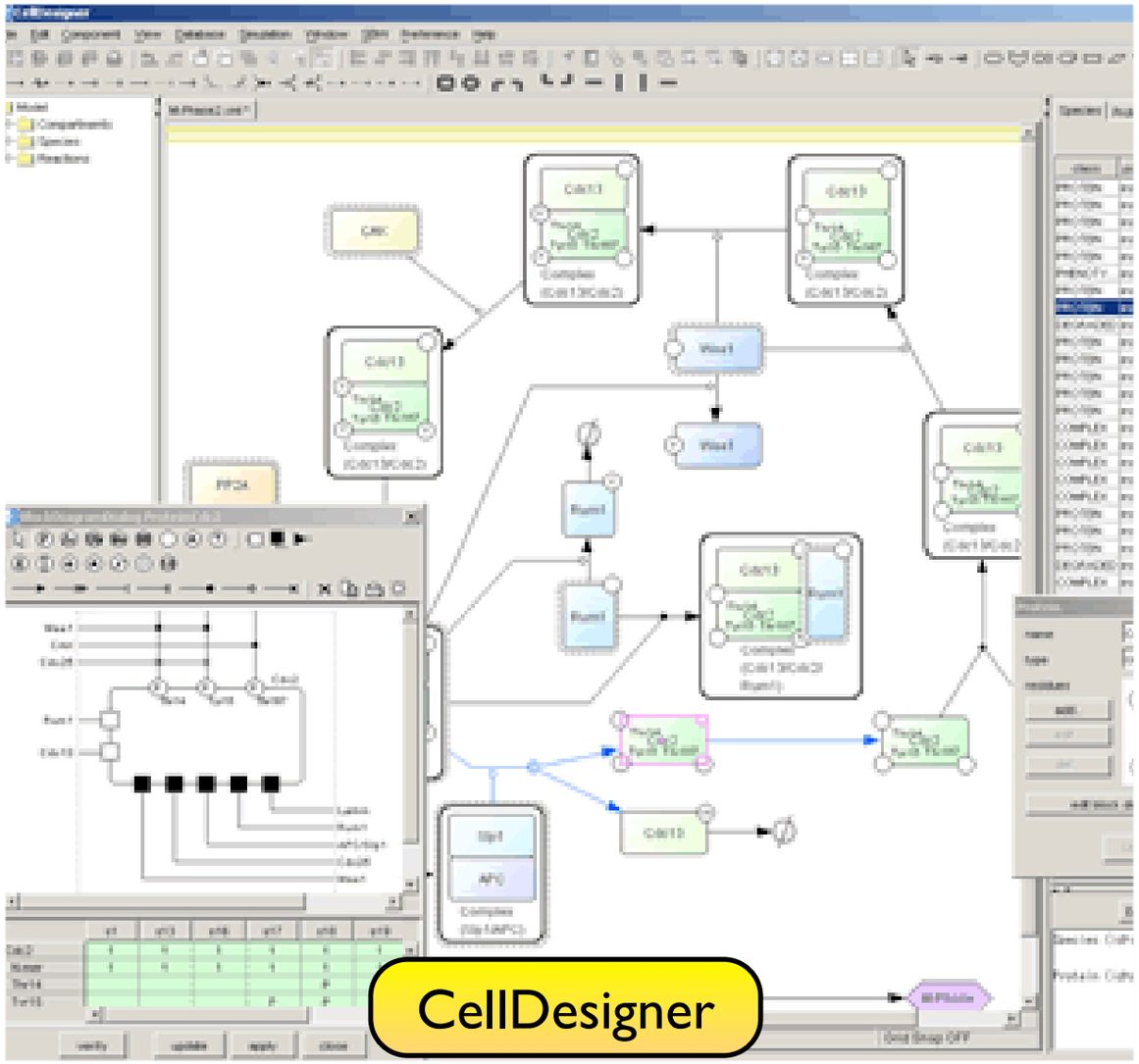
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

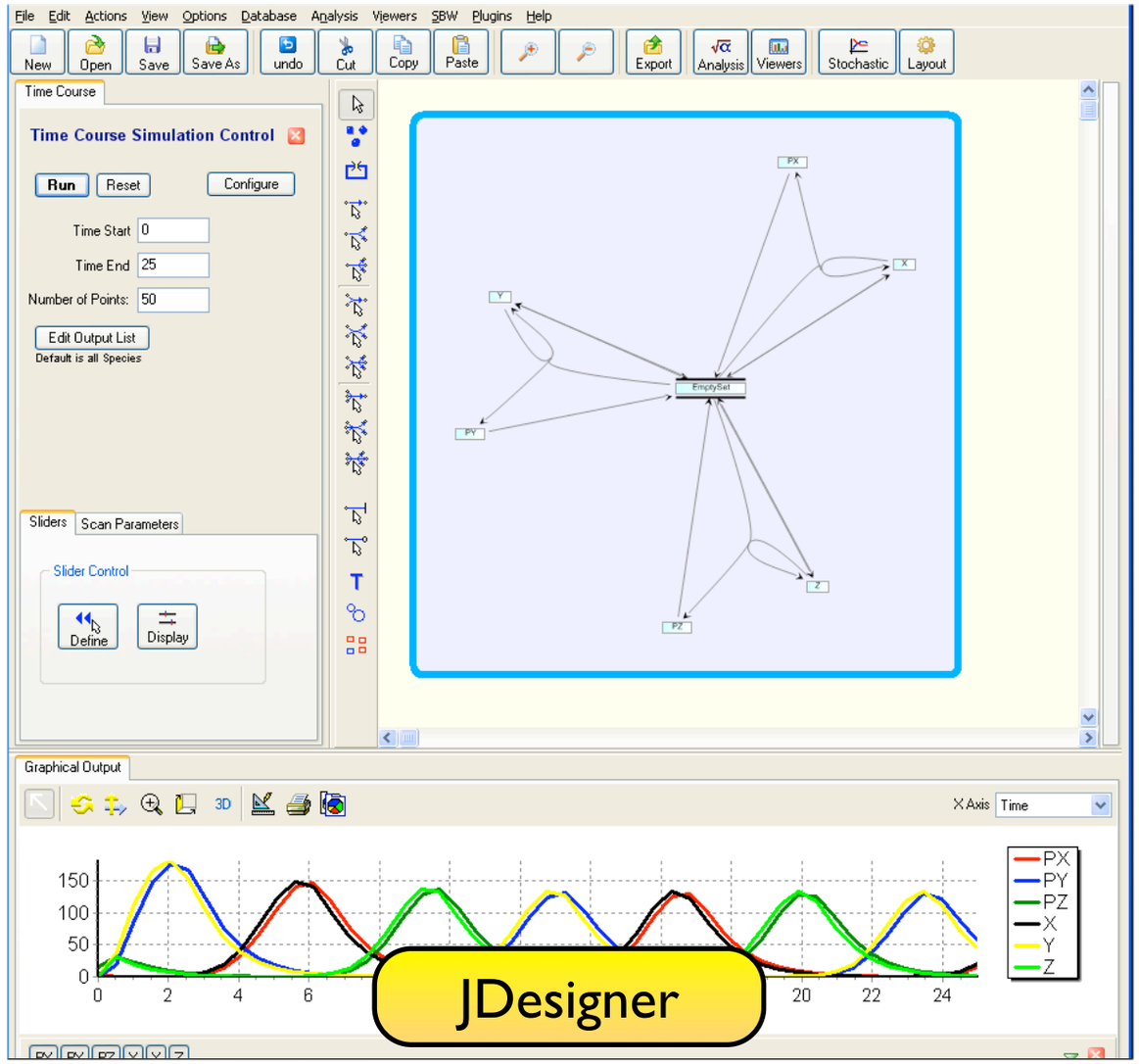
- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

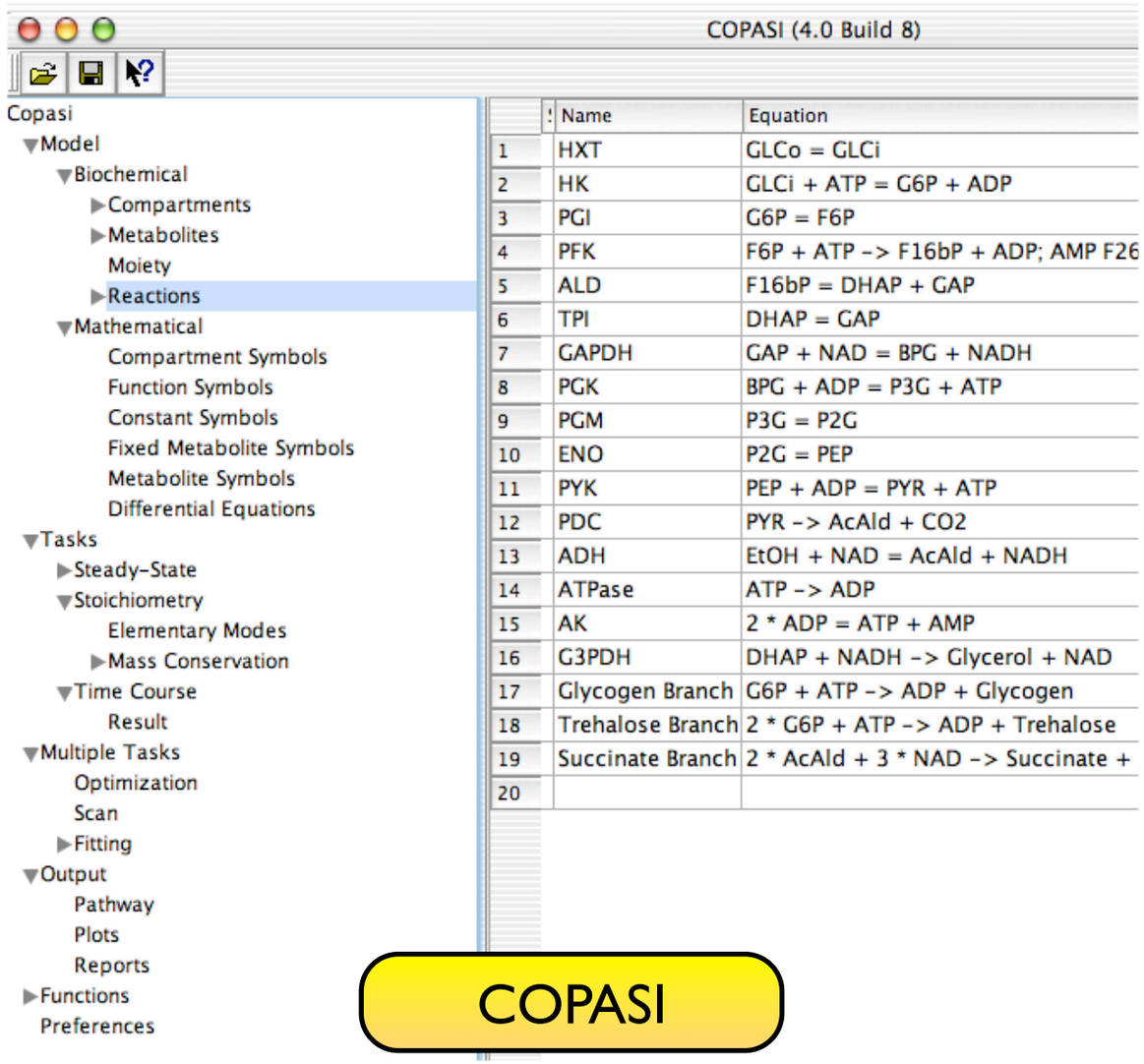
- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

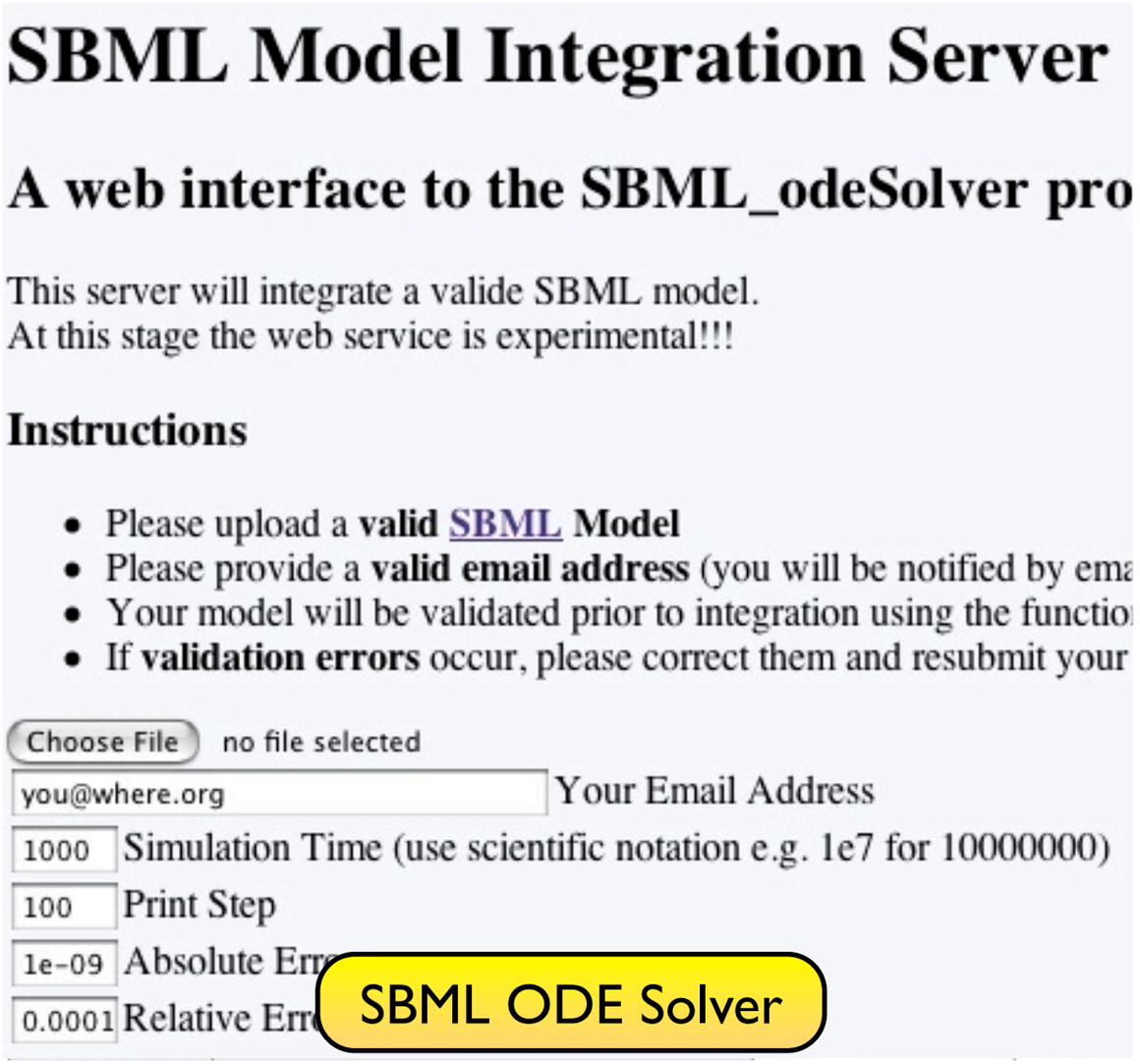
Reaction	Name	Type	
Ma->Mi	MPF inactivation	Mass Action	kw*Ma
Mi->Ma	MPF activation	Mass Action	kc*Mi
Ca->Ci	Cdc25 inactivation	Michaelis-Menten	(vcppp_
Ci->Ca	Cdc25 activation	Michaelis-Menten	(vc_*Ci*
Wa->Wi	Wee1 inactivation	Michaelis-Menten	(vw_*Wa
Wi->Wa	Wee1 activation	Michaelis-Menten	(vw_*Ww
L->	Labelled inactive MPF affected by Cdc25	Mass Action	kc*L
->L2	Labelled inactive MPF affected by Wee1	Local	kw*(1-L2
kc		Species	vcp*Ci+v
kw		Species	vwp*Wi+
vcp_		Species	vcp*Cdc
vcpp_		Species	vcpp*Cd
vcppp_		Species	vcppp/C
vwp_		Species	vwp*We
vwpp_		Species	vwpp*W
vwppp_		Species	vwppp/W
kmc_		Species	kmc/Cdc
kmcr_		Species	kmcr/Cd
kmw_		Species	kmw/We
kmwr_		Species	kmwr/W
vc_		Species	vc*Cdc2
vw_		Species	vw*Cdc2
Cdc25Total_		Species	Cdc25T

JigCell

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

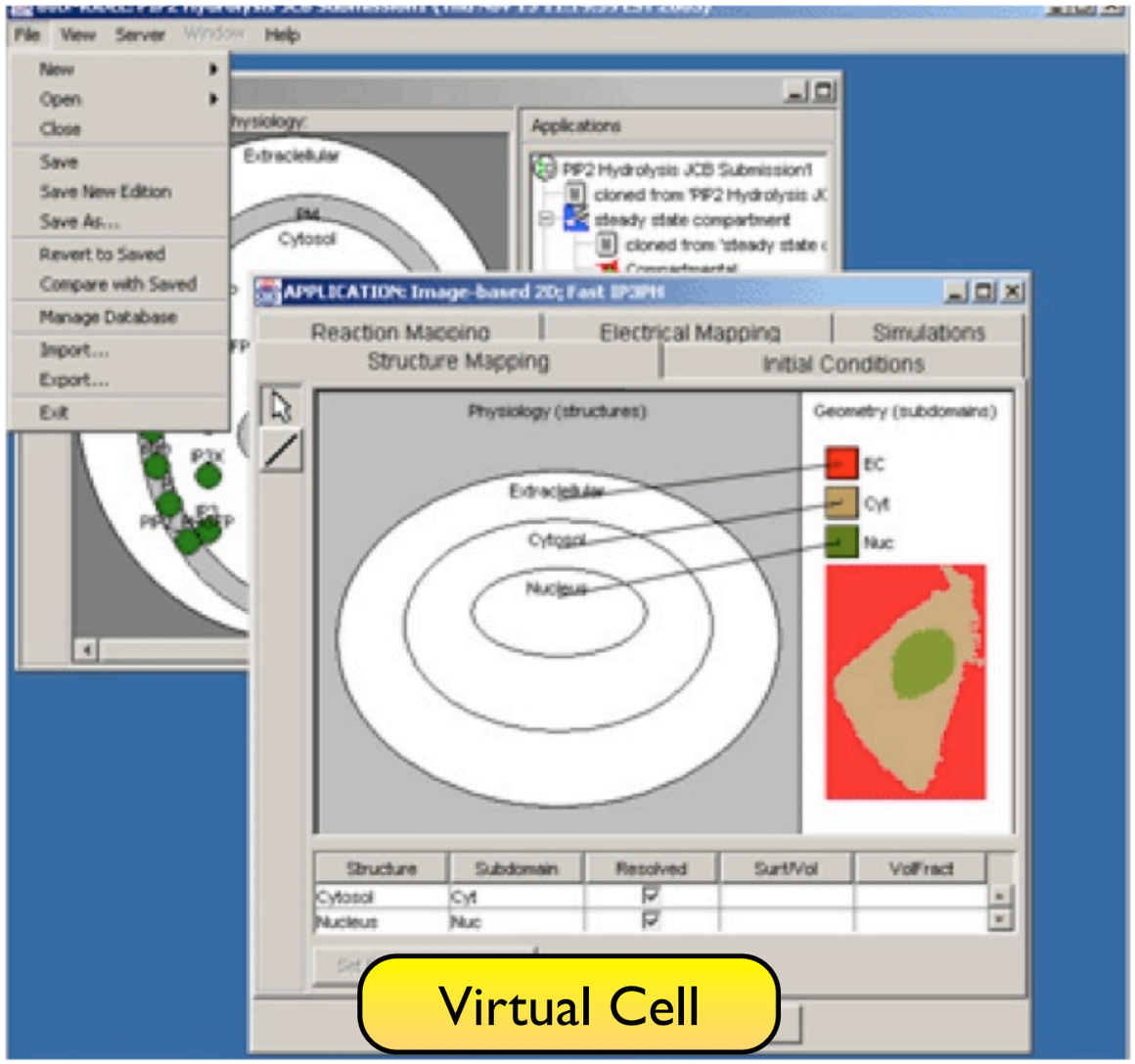
- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Many common types of model representation frameworks

- (Continuous) nonlinear differential equations
- (Discrete) stochastic systems
- Boolean networks
- Bayesian networks
- Petri Nets
- others...

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Ability to exchange models is critical

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Ability to exchange models is critical

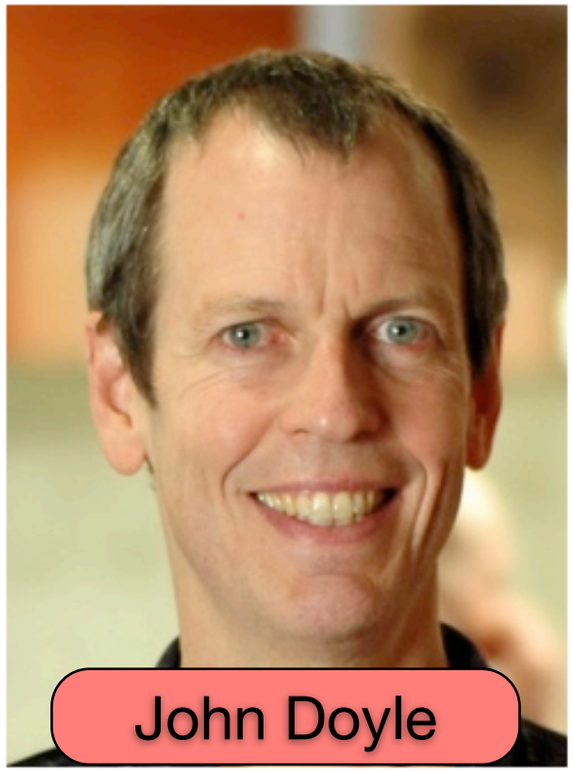
- Simply publishing equations is not enough
  - Don’t want to transcribe equations from papers
  - You want a common file format

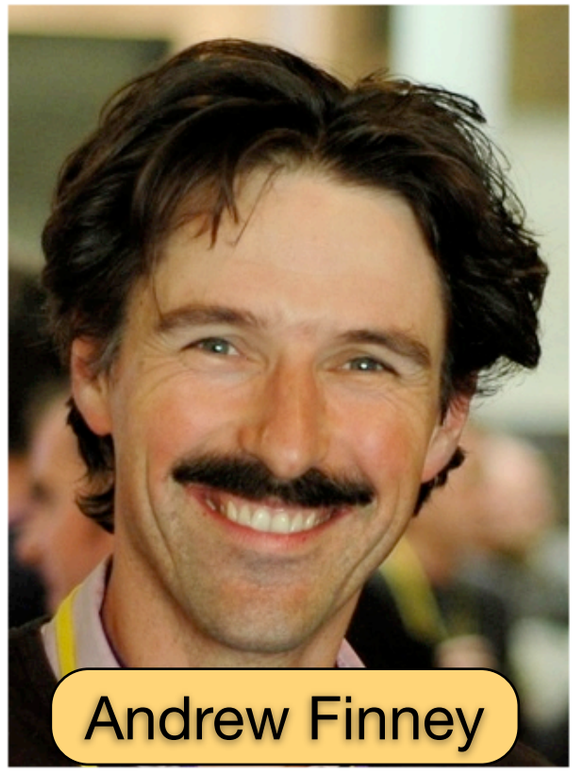
Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

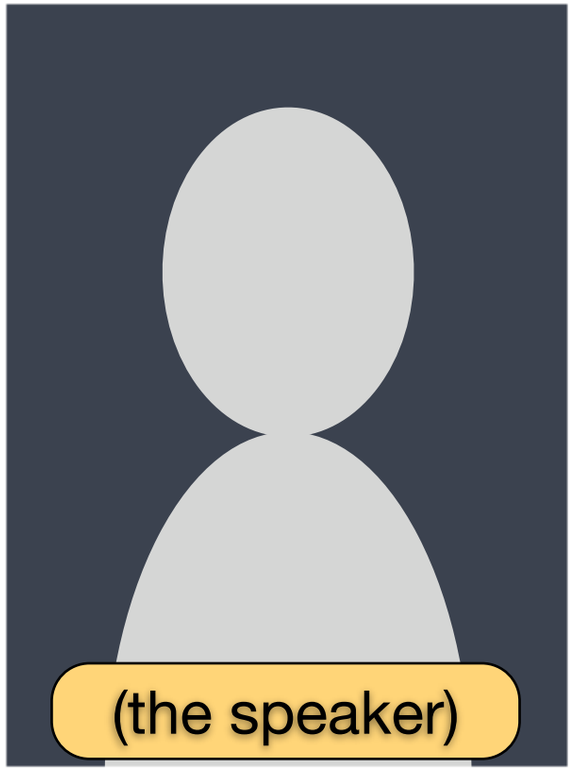
Ability to exchange models is critical

- Simply publishing equations is not enough
  - Don’t want to transcribe equations from papers
  - You want a common file format
- **Not** a new idea—seems obvious
  - Still, a format hadn’t existed before year 2000
    - Each tool had its own unique proprietary format
    - (Fewer tools too)

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007







JST ERATO Kitano Project

One initial component: get 8-10 software systems interacting

SBML =
Systems Biology Markup Language

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML = Systems Biology Markup Language

- Machine-readable format for computational models

SBML = Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
  - Arbitrary rate functions

$$2\,A + B \longrightarrow C$$

$$C \rightleftharpoons D + F$$

...

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML =
Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
  - Arbitrary rate functions
- Models can also include
  - Compartments
  - Mathematical “extras”

$$2\,A + B \longrightarrow C$$

$$C \rightleftharpoons D + F$$

...

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML = Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
  - Arbitrary rate functions
- Models can also include
  - Compartments
  - Mathematical “extras”
- Declarative, not procedural

$$2\,A + B \longrightarrow C$$

$$C \rightleftharpoons D + F$$

...

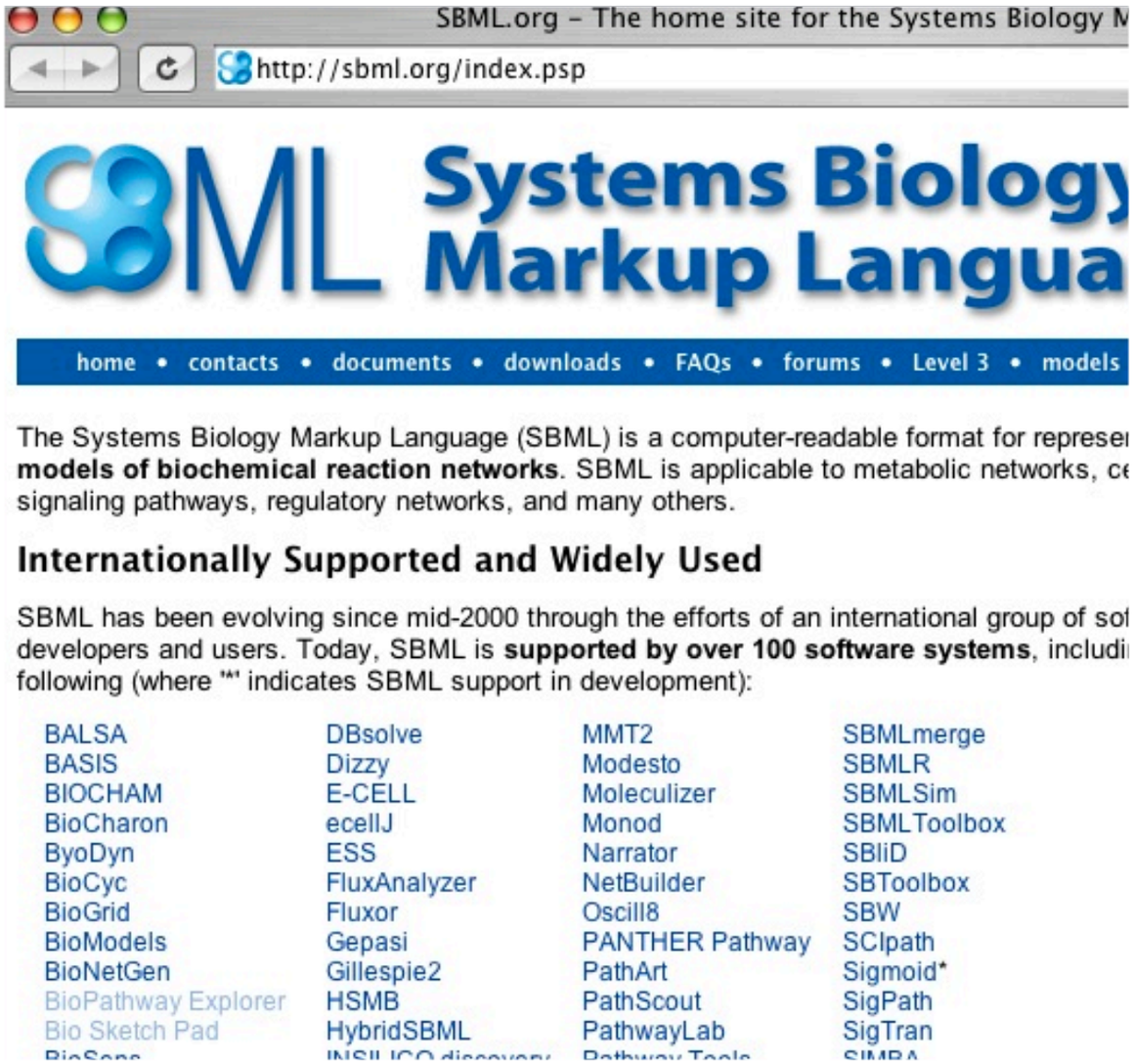
SBML is an XML format

- SBML defined using UML and XML Schema
  - Targeted at XML, but mostly independent of it
- A **lingua franca** for software, not humans
  - Think HTML

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1
      xmlns:math="http://www.w3.org/1998/Math/MathML">
    <model>
        <listOfCompartments>
            <compartment id="cell"/>
        </listOfCompartments>
        <listOfSpecies>
            <species id="X0" compartment="cell" initialConcentrati
            <species id="X1" compartment="cell" initialConcentrati
            <species id="T" compartment="cell" initialConcentratio
            <species id="S1" compartment="cell" initialConcentrati
            <species id="S2" compartment="cell" initialConcentrati
</listOfSpecies>
<listOfParameters>
    <parameter id="Keq" value="2.5"/>
</listOfParameters>
<listOfRules>
    <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
                <divide/>
                <ci> T </ci>
                <apply>
                    <plus/>
                    <cn> 1 </cn>
                    <ci> Keq </ci>
                </apply>
            </apply>
        </math>
    </assignmentRule>
```

Where is SBML today?

Now the *de facto* standard



- Supported by >100 systems
- Accepted by journals
  - Nature
  - PLoS
  - BMC
- Used in texbooks & courses

A community of modelers and software developers

- **sbml-discuss** (275+ people), **sbml-announce**
- Annual **SBML Forum** meeting (at ICSB)
- Annual **SBML Hackathon**

A community of modelers and software developers

- **sbml-discuss** (275+ people), **sbml-announce**
- Annual **SBML Forum** meeting (at ICSB)
- Annual **SBML Hackathon**

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Support by SBML Team

- Writing grants for core development
- Writing infrastructure software
  - libSBML
  - MathSBML, SBMLToolbox
- Maintaining web & mailing list resources
- Organizing workshops & other events

Latest: SBML Level 2 Version 2

- Final version released September 26

Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions

Andrew Finney
afinney@sbml.org
Physiomics PLC
Magdalen Centre
Oxford Science Park
Oxford, OX4 4GA, UK

Michael Hucka
mhucka@sbml.org
Biological Network Modeling Center
Beckman Institute, Mail Code 139-74
California Institute of Technology
Pasadena, CA 91125, USA

Nicolas Le Novère
lenov@ebi.ac.uk
European Bioinformatics Institute
Wellcome Trust Genome Campus, Hinxton
Cambridge, CB10 1SD, UK

SBML Level 2, Version 2, Revision 1
26 September 2006

Corrections and other revisions of this SBML language specification may appear over time. Notifications of revisions are broadcast on the mailing list sbml-announce@caltech.edu

The latest revision of the SBML Level 2 Version 2 specification is available at
http://sbml.org/specifications/sbml-level-2/version-2/

This revision of the SBML Level 2 Version 2 specification is available at
http://sbml.org/specifications/sbml-level-2/version-2/revision-1/

Examples of significant changes

- Many clarifications
- Simplification to the unit system
- Addition of species types, compartment types
- Addition of “constraints”
- Support for the Systems Biology Ontology (SBO)
- Recommended standard format for annotations
- “Revisions” process for handling errata

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

What lies ahead?

Revised SBML governance & development process

- Borrow ideas from W3C & other organizations
- Implement a better-defined, **regimented process**
  - Calls for proposals, etc.
  - Voting, etc.
- Have an **architectural board** to steer development
- Have **more SBML Editors**
  - Elect SBML Editors for **limited terms**

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML Level 3

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

SBML Level 3

- Modular language extensions
  - Core expected to be based “mostly” on Level 2 Version 2

SBML Level 3

- Modular language extensions
  - Core expected to be based “mostly” on Level 2 Version 2
- Layered on top of core: feature sets for—
  - Diagram storage
  - Multicomponent species
  - Models composed of submodels
  - Arrays and/or sets of components
  - Spatial geometry
  - *Other capabilities*

Full SBML Test Suite

- Allows developers to test implementation of SBML support
- Critical for improving software interoperability
- Currently have a partial “SBML semantic test suite”
- Needs further work to—
  - Complete coverage of SBML features
  - Improve ease of use
  - Update for Level 2 Version 2 and Level 3
  - Add web system for reporting results, comparisons, etc.

BioModels.net

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

Got models?

- Early FAQ in SBML: is there a database of models?

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"
  - SBML provides syntax
  - But SBML doesn't—
    - encode meaning
    - regulate use of names

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"
  - SBML provides syntax
  - But SBML doesn't—
    - encode meaning
    - regulate use of names

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1
      xmlns:math="http://www.w3.org/1998/Math/MathML">
    <model>
        <listOfCompartments>
            <compartment id="cell"/>
        </listOfCompartments>
        <listOfSpecies>
            <species id="X0" compartment="cell" initialConcentrati
            <species id="X1" compartment="cell" initialConcentrati
            <species id="T" compartment="cell" initialConcentratio
            <species id="S1" compartment="cell" initialConcentrati
            <species id="S2" compartment="cell" initialConcentrati
</listOfSpecies>
<listOfParameters>
    <parameter id="Keq" value="2.5"/>
</listOfParameters>
<listOfRules>
    <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
                <divide/>
                <ci> T </ci>
                <apply>
                    <plus/>
                    <cn> 1 </cn>
                    <ci> Keq </ci>
                </apply>
            </apply>
        </math>
    </assignmentRule>
```

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"
  - SBML provides syntax
  - But SBML doesn't—
    - encode meaning
    - regulate use of names

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1
      xmlns:math="http://www.w3.org/1998/Math/MathML">
    <model>
        <listOfCompartments>
            <compartment id="cell"/>
        </listOfCompartments>
        <listOfSpecies>
            <species id="X0" compartment="cell" initialConcentrati
            <species id="X1" compartment="cell" initialConcentrati
            <species id="T" compartment="cell" initialConcentratio
            <species id="S1" compartment="cell" initialConcentrati
            <species id="S2" compartment="cell" initialConcentrati
</listOfSpecies>
<listOfParameters>
    <parameter id="Keq" value="2.5"/>
</listOfParameters>
<listOfRules>
    <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
                <divide/>
                <ci> T </ci>
                <apply>
                    <plus/>
                    <cn> 1 </cn>
                    <ci> Keq </ci>
                </apply>
            </apply>
        </math>
    </assignmentRule>
```

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"
  - SBML provides syntax
  - But SBML doesn't—
    - encode meaning
    - regulate use of names

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/le                    ="1
      xmlns:math="http://www.w3.org/1
    <model>
        <listOfCompartments>
            <compartment id="cell"/>
        </listOfCompartments>
        <listOfSpecies>
            <species id="X0" compartment="cell" initialConcentrati
            <species id="X1" compartment="cell" initialConcentrati
            <species id="T" compartment="cell" initialConcentratio
            <species id="S1" compartment="cell" initialConcentrati
            <species id="S2" compartment="cell" initialConcentrati
</listOfSpecies>
<listOfParameters>
    <parameter id="Keq" value="2.5"/>
</listOfParameters>
<listOfRules>
    <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
                <divide/>
                <ci> T </ci>
                <apply>
                    <plus/>
                    <cn> 1 </cn>
                    <ci> Keq </ci>
                </apply>
            </apply>
        </math>
    </assignmentRule>
```

unregulated

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn't very "searchable"
- SBML provides syntax
- But SBML doesn't—
  - encode meaning
  - regulate use of names

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/le                    ="1
      xmlns:math="http://www.w3.org/1
    <model>
        <listOfCompartments>
            <compartment id="cell"/>
        </listOfCompartments>
        <listOfSpecies>
            <species id="X0" com
            <species id="X1" comp
            <species id="T" comp                    centratio
            <species id="S1" compartment="cell" initialConcentrati
            <species id="S2" compartment="cell" initialConcentrati
</listOfSpecies>
<listOfParameters>
    <parameter id="Keq" value="2.5"/>
</listOfParameters>
<listOfRules>
    <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
                <divide/>
                <ci> T </ci>
                <apply>
                    <plus/>
                    <cn> 1 </cn>
                    <ci> Keq </ci>
                </apply>
            </apply>
        </math>
    </assignmentRule>
```

unregulated

low info content

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

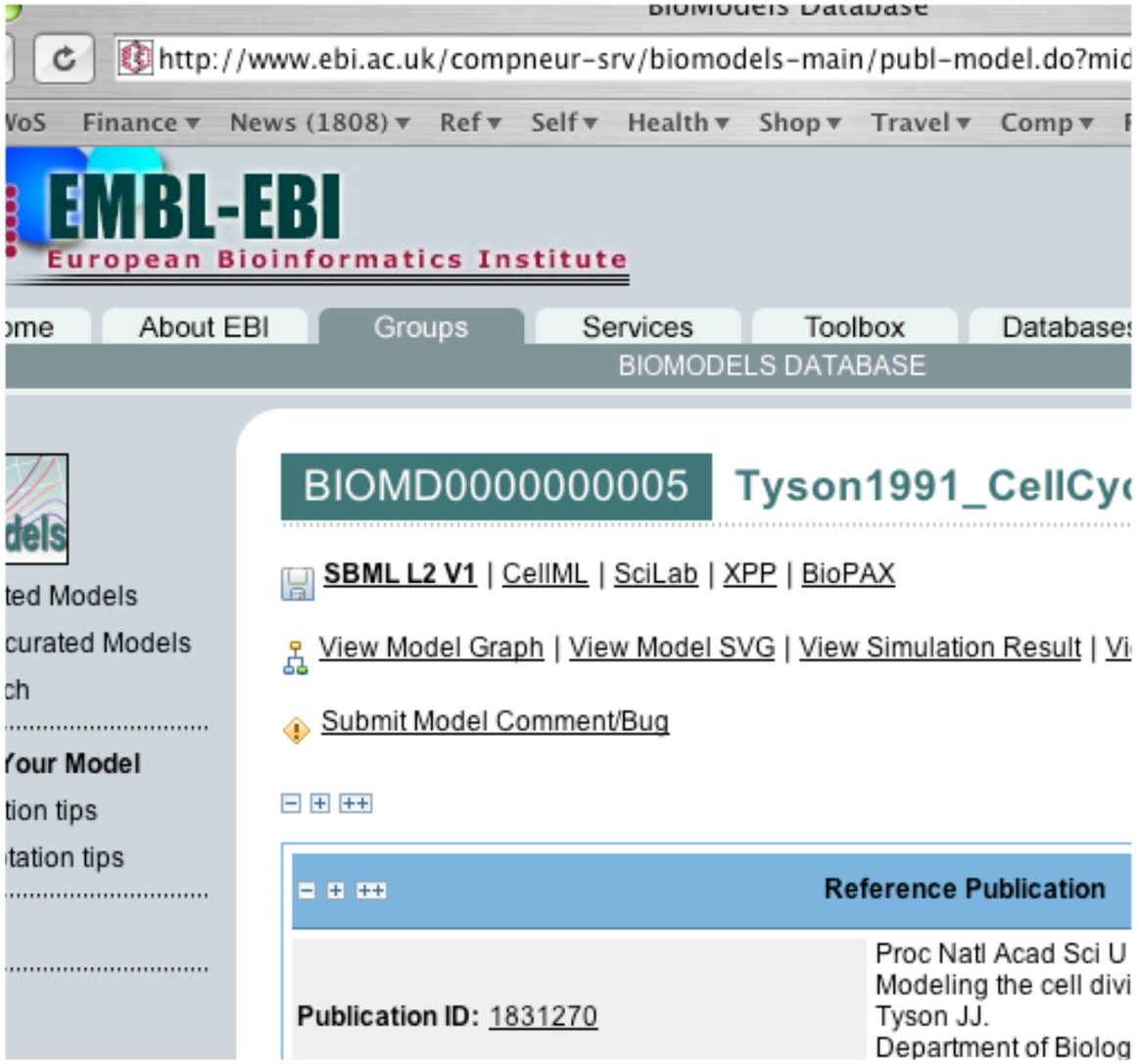
BioModels.net consortium

- International collaboration to develop:
  1. A public database of **curated**, **annotated** models: **BioModels Database**
  2. Guidelines for curation and annotation of models: **MIRIAM**
  3. Ontology suited to computational models: **SBO**
- Main collaborators
  - Nicolas Le Novère's Computational Neurobiology group at EBI (UK)
  - Hucka and SBML Team at Caltech (USA) and U. Hertfordshire (UK)
  - Herbert Sauro's group at Keck Graduate Institute
  - Hans Westerhoff & Jacky Snoep's JWS Online (ZA and UK)
  - Hiroaki Kitano's Systems Biology Institute

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

BioModels Database
http://www.ebi.ac.uk/biomodels

- Stores & serves **quantitative** models of bio. interest
- **Free, public** resource
- Models must be described in **peer-reviewed** publication(s)
- Imports models in **SBML** & **CellML** formats
- Exports in **SBML**, **CellML**, **SciLab**, **XPP** and **BioPAX**



Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Models are curated

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Models are curated

- Human curators check correspondence to publication

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Models are curated

- Human curators check correspondence to publication
- Human curators add annotations

Models are curated

- Human curators check correspondence to publication
- Human curators add annotations
  - General info about model, author, publication, etc.

Models are curated

- Human curators check correspondence to publication
- Human curators add annotations
  - General info about model, author, publication, etc.
  - References to external data resources: UniProt, KEGG, NCBI, Gene Ontology, ChEBI, BIND, Reactome

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Features of BioModels Database

- **True database**: can search model content & annotations
  - SBML XML stored in XINDICE; annotations in an SQL database
- Some statistics today (6th release in Oct. 2006):
  - **70+ curated, dynamical models**
    - E.g.: Tyson yeast cell cycle models, Elowitz E. coli repressilator, Teusink et al. yeast glycolysis, Rohwer et al. E. coli glucose transport
  - 26 non-curated models (e.g., models lacking kinetics, such as FBA)
- Model sources: us, Nature/EMBO ***Molecular Systems Biology***, repositories such as JWS Online and CellML, individual researchers

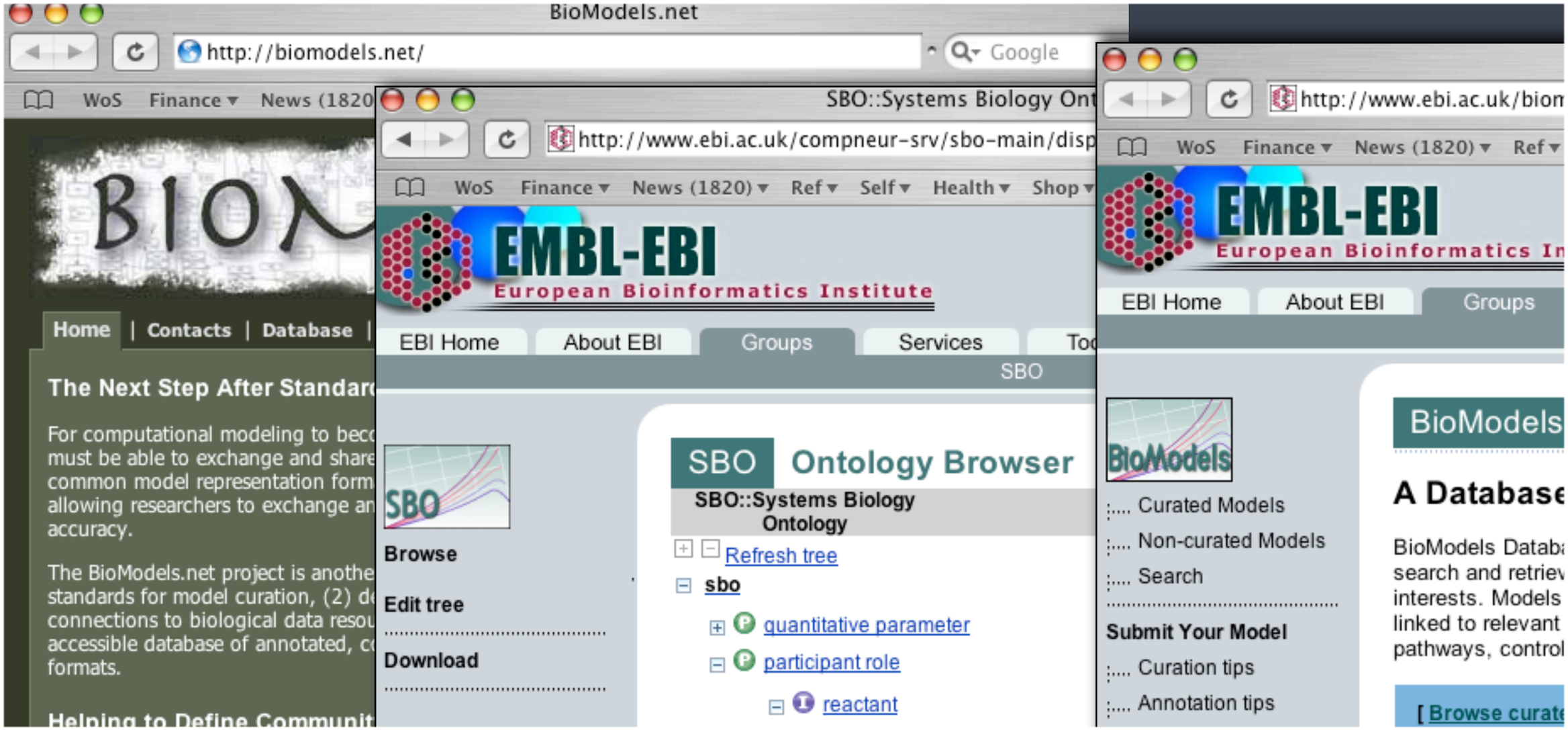
MIRIAM

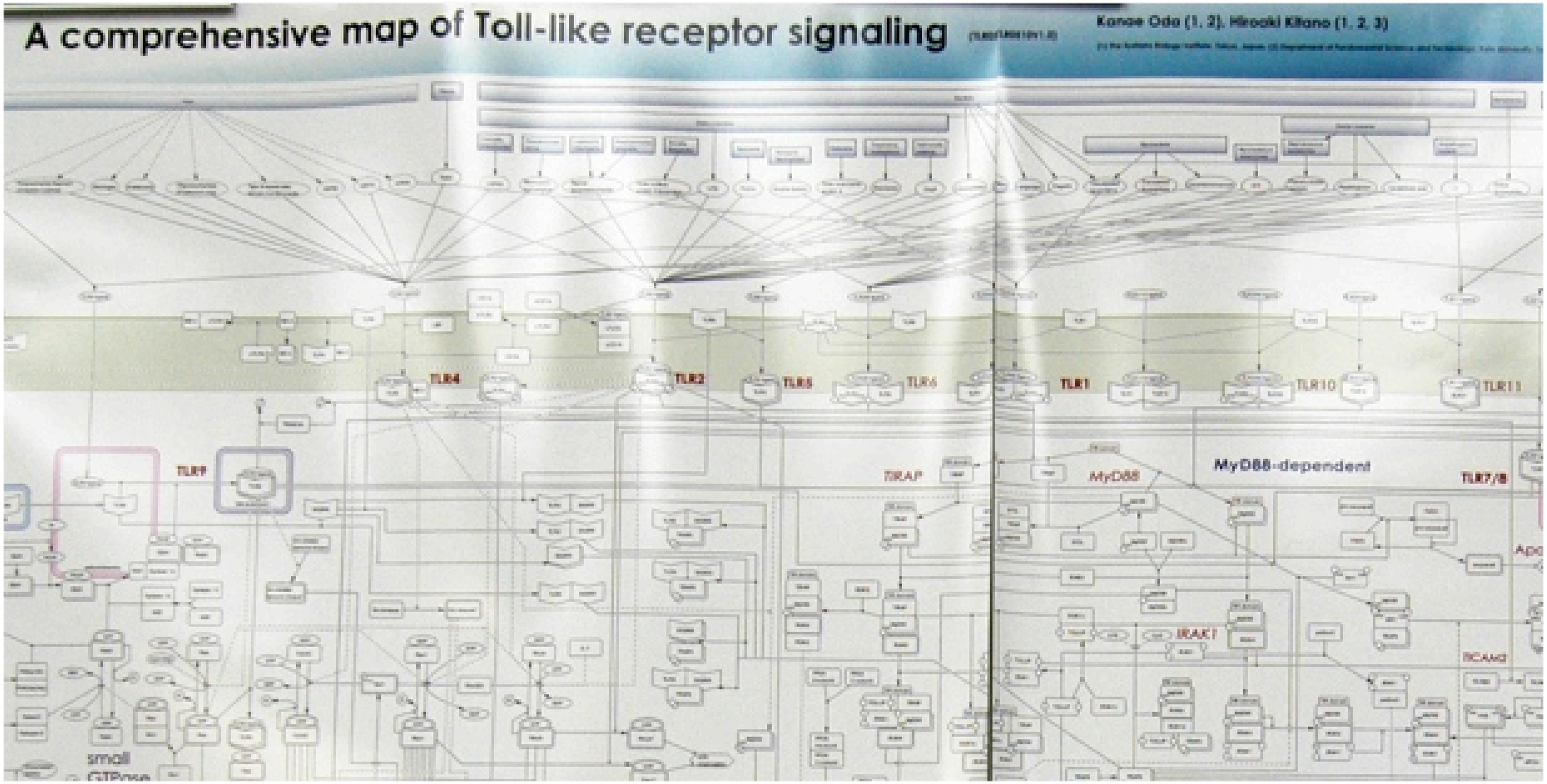
- "Minimal info. requested in the annotation of biochemical models"
- Proposed guidelines for basic annotation of models
  - **Reference correspondence**, e.g.,
    - Be encoded in a public, standardized format (SBML, CellML, etc.)
    - Must be instantiated in a simulation & all quantitative attributes defined
  - **Attribution annotation** (info about model creators, source reference, etc.)
  - **External source annotations** (linking model element to data source)
- Goal: minimal common standards enabling sharing of curation effort

SBO = Systems Biology Ontology

- Occupies a space not filled by other ontologies
  - Primarily for describing rate laws and constituents
    - **Classification of rate laws**
      - Each term includes a mathematical function definition
    - Controlled vocabulary for the **roles of reaction participants**
      - E.g.: “substrate”, “catalyst”, “competitive inhibitor”, etc.
    - Controlled vocabulary for the **roles of parameters** in quantitative models

BioModels.net portal




A comprehensive map of Toll-like receptor signaling
Kanae Oda (1, 2), Hiroaki Kitano (1, 2, 3)
TLR4
TLR2
TLR5
TLR6
TLR1
TLR10
TLR11
TLR9
TIRAP
MyD88
MyD88-dependent
TLR7/8
IRAK1
small

SBGN

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

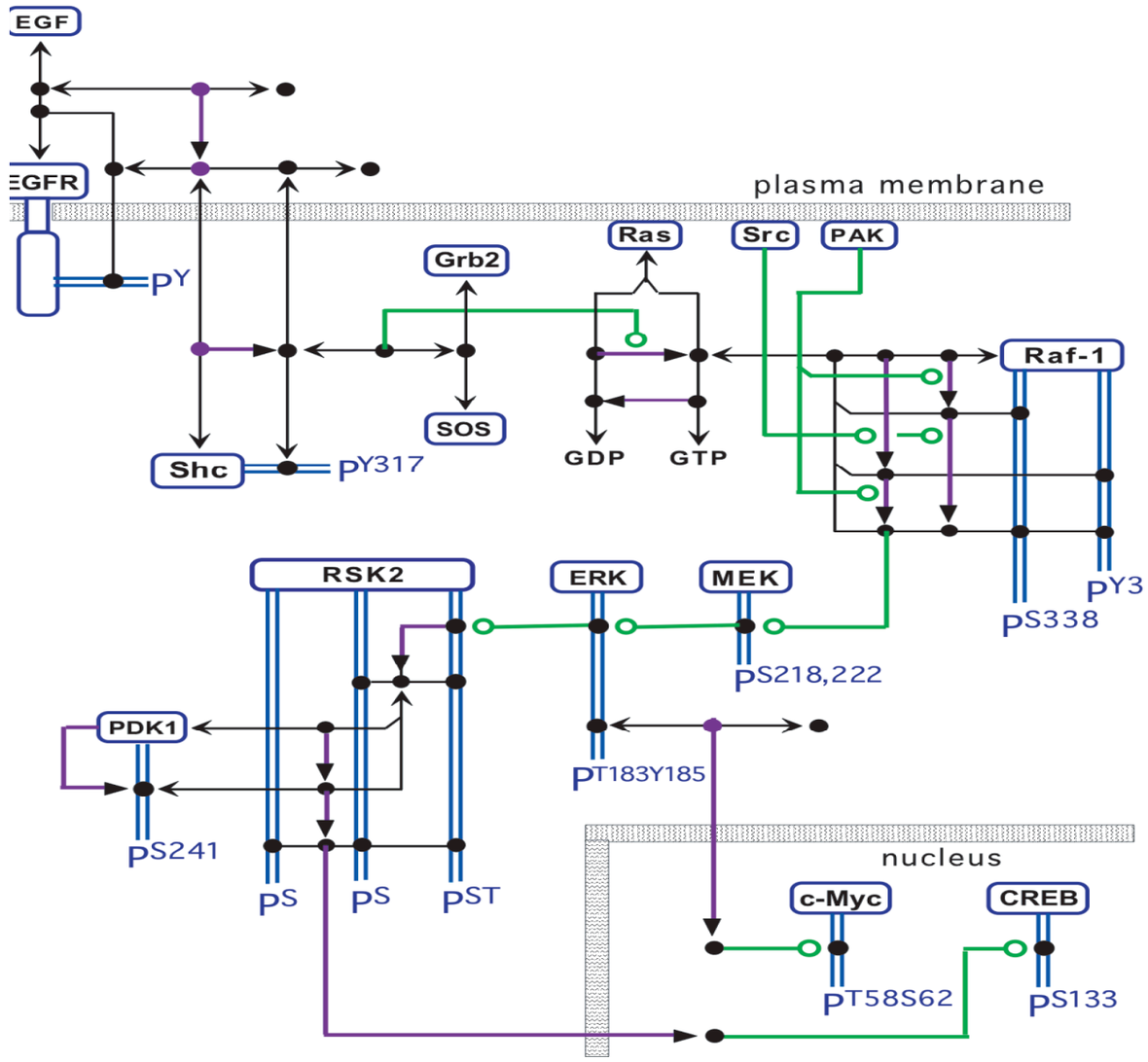
Background

- No current standard for network diagrams in biology
- No consistency—
  - Between authors
  - Between papers
  - Between publications

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Background

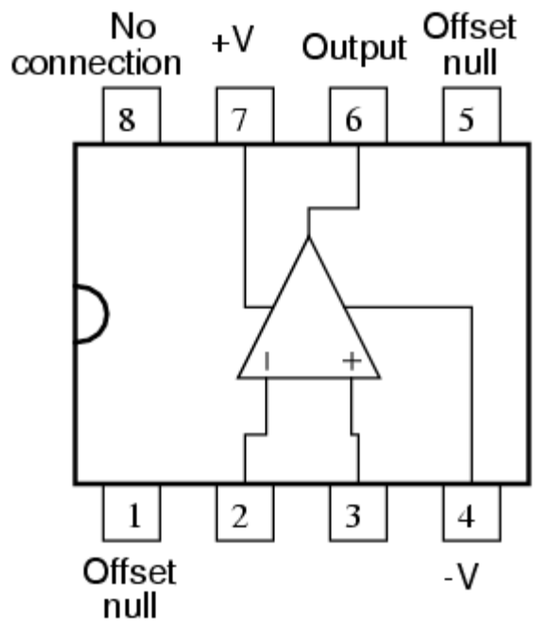
- No current standard for network diagrams in biology
- No consistency—
  - Between authors
  - Between papers
  - Between publications



Standardization would bring benefits

- Notations standardized in electrical/electronics, software engineering, etc.
- Taught in textbooks
- Supported by software
  - Automated verification
- Consistency makes it easier to read new diagrams

Typical 8-pin "DIP" op-amp integrated circuit



SBGN = Systems Biology Graphical Notation

- Goal: bring simulation/modeling community together and develop a proposal for a standard notation for **some** types of diagrams
  - Starting with metabolic & signaling networks
- Begun late ‘05 by Kitano/Le Novère/Hucka thanks to NEDO funding
- 2 SBGN workshops held so far (Feb and Oct 2006)
  - Many groups participating: Goryanin group (U. Edinburgh), BioPAX (MSKCC New York), SRI, CellML, EML (Germany), many others
  - Join if you’re interested!
- Currently working towards a first proposal

Closing

The funding

- National Institute of General Medical Sciences (USA)
- JST ERATO Kitano Symbiotic Systems Project (Japan) (to 2003)
- National Science Foundation (USA)
- International Joint Research Program of NEDO (Japan)
- JST ERATO-SORST Program (Japan)
- Japanese Ministry of Agriculture
- Japanese Ministry of Educ., Culture, Sports, Science and Tech.
- BBSRC e-Science Initiative (UK)
- DARPA IPTO Bio-SPICE Bio-Computation Program (USA)
- Air Force Office of Scientific Research (USA)
- STRI, University of Hertfordshire (UK)
- Beckman Institute, Caltech (USA)

SBML Team	BioModels DB Team	SBGN Team
Michael Hucka	**Nicolas Le Novère**	**Hiroaki Kitano**
Andrew Finney	**Chen Li**	**Nicolas Le Novère**
Ben Bornstein	**Mélanie Courtot**	**Michael Hucka**
Sarah Keating	**Lu Li**	**Akira Funahashi**
Bruce Shapiro	**Camille Laibe**	
Ben Kovitz	**Nicolas Rodriguez**	
Hamid Bolouri	**Harish Dharuri**	
Herbert Sauro	Marco Donizelli	
Maria Schilstra	Alexander Broicher	
Jo Matthews	Arnaud Henry	

A million thanks to the SBML Community too

Nature Precedings : doi:10.1038/npre.2007.20.1 : Posted 22 Jan 2007

Where to learn more

- **http://sbml.org**
- **http://biomodels.net**
- **http://sbgn.org**
- *Upcoming*:
  - BioModels.net Training Camp 2007 January 13-15, Manchester, UK
  - SBML Hackathon 2007 in June at U. Newcastle, UK
  - SBGN Workshop in 2007 (probably March 2007, Gosau, Austria)
  - SBML Forum 2007 in Long Beach, CA, USA, Oct. 5-6 (ICSB 2007)
- **Thank you**!